GUIDED PROJECT: SIMPLE LINEAR REGRESSION FOR THE ABSOLUTE BEGINNER

By Dr. Ryan Ahmed

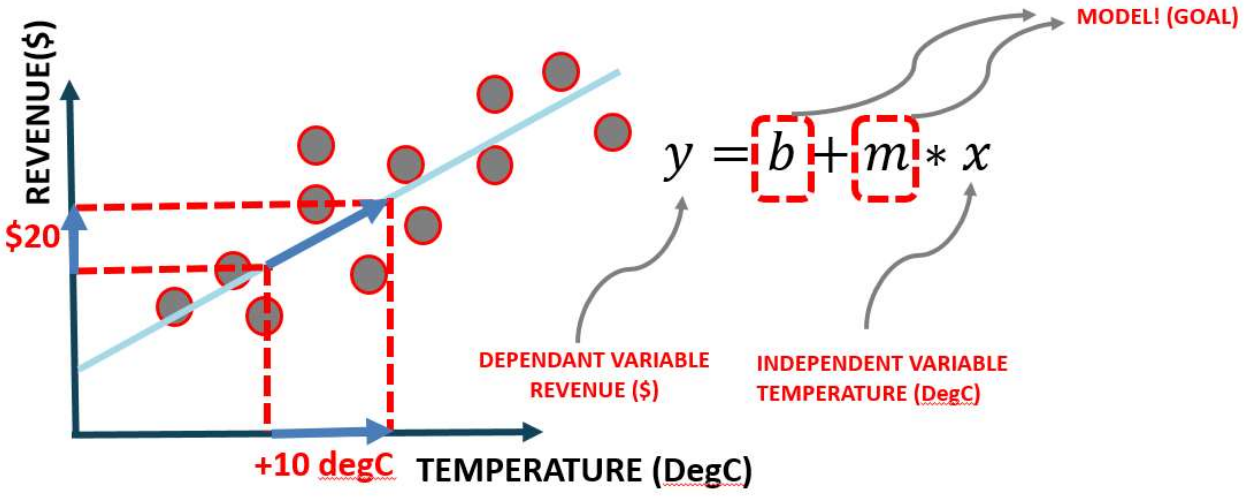
Hello everyone and welcome to this hands-on guided project on simple linear regression for the absolute beginner. In simple linear regression, we predict the value of one variable Y based on another variable X. X is called the independent variable and Y is called the dependent variable.

1. KEY LEARNING OUTCOMES:

1. Perform data cleaning, feature engineering and visualization
2. Build, train and test a simple linear regression model in Scikit-Learn library
3. Understand the theory and intuition behind simple linear regression models

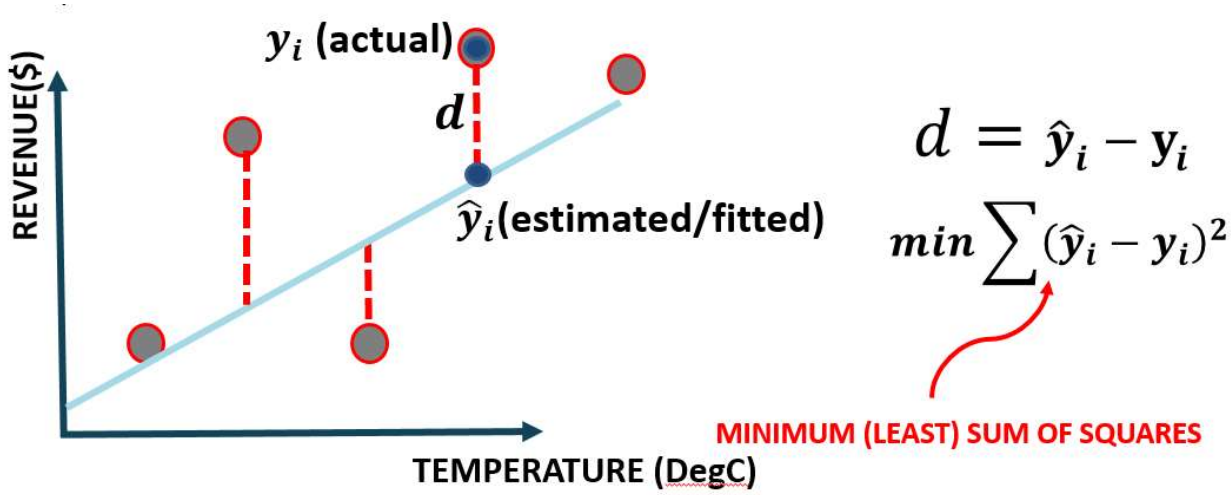
2. SIMPLE LINEAR REGRESSION

- Simple linear regression is used to predict the value of one variable Y based on another variable X.
- X is called the independent variable and Y is called the dependant variable.
- Why do we call it simple? Because it examines relationship between two variables only.
- Why do we call it linear? Because when the independent variable increases (or decreases), the dependent variable increases (or decreases) in a linear fashion.



3. LEAST SUM OF SQUARES

- Least squares fitting is a way to find the best fit curve or line for a set of points.
- The sum of the squares of the offsets (residuals) are used to estimate the best fit curve or line.
- Least squares method is used to obtain the coefficients m and b.



4. BUILD A SIMPLE LINEAR REGRESSION MODEL USING SCIKIT-LEARN

```
>> from sklearn.linear_model import LinearRegression
>> regressor = LinearRegression(fit_intercept = True)
>> regressor.fit(X_train,y_train)
>> print('Linear Model Coefficient (m): ', regressor.coef_)
>> print('Linear Model Coefficient (b): ', regressor.intercept_)
```

5. EVALUATE THE MODEL (MAKE PREDICTIONS USING TRAINED MODEL)

```
>> y_predict = regressor.predict( X_test)
>> y_predict
```

6. DIVIDE DATASETS INTO TRAINING AND TESTING

Data set is generally divided into 75% for training and 25% for testing.

- Training set: used for model training.
- Testing set: used for testing trained model.

Make sure that testing dataset has never been seen by the trained model before.

7. DIVIDING DATASET USING SCIKIT-LEARN

```
>> from import   sklearn.model_selection train_test_split
>> X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=0)
```

Final Capstone Project Materials Link:

https://drive.google.com/drive/u/0/folders/1FuyDhsGfz5ImbgT29_NsR0haJZzI20zB